ISSN: 0976-9102 (ONLINE)
DOI: 10.21917/ijivp.2020.0324

MULTIPLE ATTRIBUTE FEATURE EXTRACTION AND HIGH SUPPORT VECTOR CLASSIFIER FOR IDENTIFICATION OF CYTOMEGALOVIRUS IMAGES

K. Deepa and S. Suganya
Department of Computer Science, Rathnavel Subramaniam College of Arts and Science, India

Abstract

Digital image processing is a broad area consists of different concepts to emphasize the quality of the images. The digital image processing has been engaged in various areas such as remote sensing, video processing, image sharpening, colour, pattern recognition, image compression etc. The digital image processing plays an immeasurable role in the field of biomedical. The clarity of the images are important for any types of medical diagnosis leads with human body. In this paper the image of cytomegalovirus is taken and processed to show the improvement of the affected images. Though there are many facilities to identify any type of diagnosis, detection of viruses in terms of images also plays a vital role. The cytomegalovirus affects almost all ages of people, but it shows more impact on pregnant women, if the women gets infected the foetus also gets affected so this leads to congenital infections of the baby such as impair vision, autism, hearing loss, birthmarks etc. Once the person gets affected it retains for lifetime. In this proposed work the input of the image is taken and processed with ROI and MAFE Algorithm with PCA reduction and HSVM is used for classification and Identification of the image is done.

Keywords:

Cytomegalo Virus, Image Processing, Feature Extraction

1. INTRODUCTION

The human body has to undergo many changes from birth to death. The survival of the fittest deals with the immunity system of the human body. Due to many environment changes and pollution lots of difficulties are faced by living organisms. Even though there are n number of viruses few viruses' shows more impact on humans. In order to that cytomegalovirus is one among them. The cytomegalovirus once entered in the human body stays for lifetime sometimes. It shows more trouble to the pregnant women. Once the pregnant women gets affected, the fetus also gets affected which leads to more complications during the birth as well as after birth. The impact of the virus cannot be seen during the time of scan. The congenital cytomegalovirus means a condition where a virus that is occurred during the time of birth to an infant is affected before birth or immediate after birth. This Cytomegalovirus infection occurs when they transmit using saliva, semen, body fluids, sexual transmission, blood vessels, etc. The virus doesn't show any symptoms because the virus remains in the body if the person gets affected for the first time and it retains in the body till lifetime as the study shows about the virus it is there inside the human body for the lifetime if the person has a very weak immune system, then the virus gets reactivated and starts its impact on the human body. The most serious issue among the pregnant women is that the women gets pregnant and the dream of every pregnant women is to deliver a healthy baby. The pregnant women has to undergo hormonal changes so most of the research says that pregnant women are more sensitive among certain infections and one among them is cytomegalovirus infections. The human body has five types of immunoglobulins (antibodies)-IgM, IgG, IgA, IgD and IgE each has its functionality to build the immune system of the human body the one among them is IgG which the antibody which carries all types of nutrients to the womb for the fetus through the placenta. If the pregnant women is affected by the CMV virus then the IgG antibody carries the infection also through the placenta. So the fetus gets affected and it doesn't show any symptoms until the infant is born. If the human body improves the immune system then slowly the impact of the virus gets decreased. Somehow if the immune system gets weakened then the virus gets reactivated and shows the symptoms. The most problems occurs in infants are developing hearing loss, impaired vision, autism, birthmarks, etc. As all come to know about the medical tests (thyroid tests, diabetic tests) are taken when the women gets pregnant these are made mandatory to all the women because if there is any complications in the above tests then there might be more problems which will occur to the fetus. But all these years cytomegalovirus is not considered so far but around the globe there are people who got affected about all ages. 1 to 5% of birth has some non-primary maternal infections worldwide.

Mostly CMV can cause mononucleosis or hepatitis mainly the liver problem. It can also affect the eyes, ears, intestines. Babies born with CMV infections have the growth related problems. In the field of medicine there are so many complications in the human body and most of the diagnosis is done through the images of the parts of the human body by CT scan, X-ray Images, MRI images, microscopic images etc. The more detailed information can be performed using the digital image processing and it plays a tremendous role in all fields and it renders more scintillating role in medical field. The input image of CMV is depicted and it represents like an owl's eye. The various types of images of cytomegalovirus is portrayed and different views taken from the body fluids, blood vessels, saliva. In this proposed system the image of cytomegalovirus is taken and processed under various algorithms to form a system which deals with the classification of fully affected, partially affected and uninfected will be differentiated for identification of the virus. The ROI segmented image is taken and processed using the Multiple Attribute Feature Extraction with PCA reduction and support Vector machine is used for classification and Identification of the virus image is done.

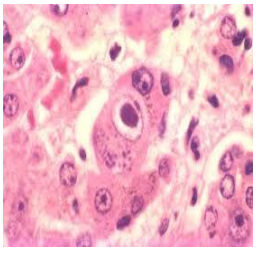

Fig.1. Input Image of Cytomegalovirus

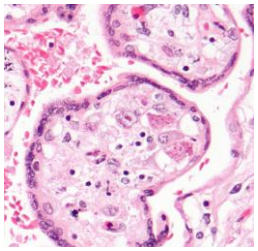

Fig.2. Input Image of cytomegalovirus with a different view

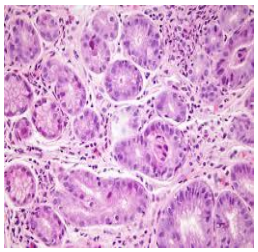

Fig.3. Images of Cytomegalovirus with the body fluids.

2. RELATED WORK

The cytomegalovirus is identified by medical tests such as PCR Polymerase Chain Reaction on saliva and urine which is collected and tested for the confirmation of the virus. The terminology used in mothers are called as seropositive mothers where they shed CMV in their breast milk. If there is a suppression in the immune system then the virus can get reactivated.

The enzyme-Linked immunosorbent assay (ELISA) it is the test to measuring the antibodies to CMV it is a serological test. The samples are taken and tested if the IgG is negative and in the second sample if the infection is identified then it leads to recent infections that is the person has got affected recently and it leads to recent primary infection. Digital image processing has an impact in the field of biomedical. The identification of affected images are done with the help of the algorithms used in image processing. The various types of identifications are done in the case of cytomegalovirus using the decision tree analysis that can be done as rare event cytometry. The count of cells shows the variations in terms of cell difference. But one need to be monitored regularly.

The Polymerase Chain Reaction (PCR) is used as a method of diagnose cytomegalovirus. It has the qualitative and Quantitative determination of CMV and shows the image quality assessment. Most of the digital image processing algorithms are used to detect the clarity of the virus detection and the impact is shown, so that it is clear to identify with most accuracy. The congenital infections are the main cause for most of the trauma in human body [1] [20]. Neonatal care is important and identification of neonatal problems can be solved at the earliest [18]. Epilepsy also can be caused by CMV infections [19]. Mostly the images are framed in different types of sources like x-ray, CT scan, ultrasound etc. In all these sources images plays a vital role and the accuracy of the detection is more compared to any other sources. Here in this proposed work the image of cytomegalovirus image is processed under various algorithms to show much clarity and the system is trained and tested to show the performance analysis based on affected, partially affected, unaffected.

3. METHODOLOGY

The methodology depicts here is the process of the cytomegalovirus image that is processed under the various forms such as hierarchical ranking convolutional neural network and image background subtraction to make the image as a more intensified image the image enhancement is done and extracted as the intensified image which improves the image detail for further processing. The CMV image is taken as Input and the image is processed and Segmented under the K-Means Clustering with Discrete Wavelet Transform. With the Region of interest, the image is further processed under the feature extraction of MAFE Algorithm with PCA reduction to show better performance. Using the HSVM the collection of images are trained under group train and tested based on support vector machine to classify and the identification of the Image is done.

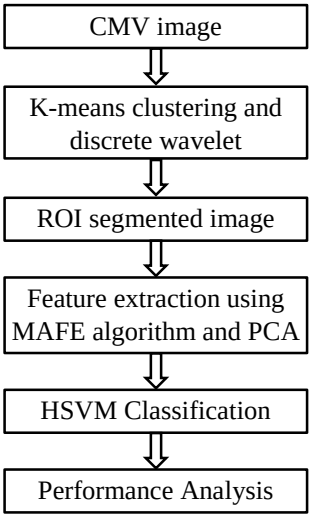

Fig.4. Block Diagram of Proposed System

The main aim of this proposed system is to interpret the performance of the image under various functionalities to show the better clarity of the images. The block diagram depicts the representation of the proposed system in determining the more specific and concise information analysis about the system. In the previous work the image is processed under the clustering phase and wavelet forms, the next to proceed with the process is to find out the more detailed information about the performance analysis by using the algorithm (MAFE-Multiple Attribute Feature Extraction algorithm using the Principle Component Analysis. Classification using the High Support Vector Machine.

The CMV(Cytomegalovirus) image is taken and processed under the image enhancement techniques by using the convolutional neural network where the 15 layers have been processed forward and backward and by using the image background subtraction to find out the observation of images under the microscopic view and to identify the details of the images even under the atmospheric light estimation to bring out the intensity levels of the image to find out each and every single detail of the image without the loss of information. The K-means clustering and discrete wavelet transform is used to determine the cluster index of the images. Segmenting the images will ensemble the preferred information and to find out the features of the input

CMV images. The discrete wavelet transform is used to provide the multi-resolution shaped on time-scale representation. The main objective of the discrete wavelet transform is to provide denoise two dimensional signals such as images. The images are reconstructed based on various modified levels filters, etc.

3.1 REGION OF INTEREST (ROI)

ROI detection has been studied for many years. The methodology used to find out the boundaries of each and every specific detail and to know about the cytomegalovirus affected areas defined on the image of the blood samples, body fluids, semen and saliva. It clearly defines the boundaries of an object on a particular images. ROI determines the points of interest denote the region or a volume with specific amenities. Image segmentation plays an important role and deals with direct influence is made and performs the results where the extraction is done in the images and shows more specific detail for further clarification.

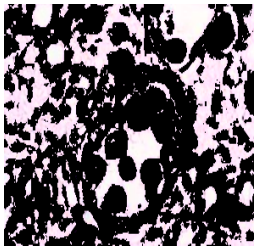

Fig.5. Segmented ROI of Cytomegalovirus

The general use or a common use of ROI is to create a binary mask where the pixels are set to 1 if the pixels are among the Region of interest and outside the Region of interest is set to 0. The ROI is a part or portion of an image that one wants to filter and operate to know how it is represented and the functioning of the methodology to show more appropriate value to make the image more concise Most algorithms use either feature-based or object-based approaches. Feature-based methods find pixels that share significant optical features with the target and aggregate them to form ROI. These methods can capture most of the target pixels on the basis of optical feature similarity. However, not all target pixels have strong optical features, so the detected ROI usually fails to encompass the entire target. In addition, feature-based methods cannot distinguish between targets, which can cause confusion in subsequent stages of processing. The significance of the image improves the performance of the image classification.

3.2 MAFE ALGORITHM

In this proposed work the Multiple Attribute Feature Extraction (MAFE) algorithm is used for the clarity of the images based on Wavelet, statistical, textural and Geometrical attributes which is a hybrid where each one depicts its own functionality to make the images more precise. Research have been made under various circumstances to make more clear and resourceful to ensure and detect most of the problems and finds out an assortment of types of solutions to form a improved one. The ultimate resist of findings the solutions is to reach out better results.

3.2.1 Texture Attribute:

The classification or segmentation of images can be done using many ways, among those image texture plays an immeasurable role, it deals with the spatial frequency and an average grey level [14]. It deals with the measures of intensity, color or region based with convention of the occurrence of an image [15]. Every single information is noted to quantify the perceived texture of an image. The characterization of texture analysis deals with the surface of the given object [2]. The overall texture analysis is to perform the unique identity of each and every pair of pixels [3]. Texture parameters are used for image classification [4]. It is all about texture feature extraction. The texture are of four type's actual, simulated, abstract and invented texture. Texture classification has the strength and weakness based upon the techniques which is been used. The features of Grey-Level Co-occurrence Matrix (GLCM) which calculates the specific value with pair of pixels and spatial relationship occurs in an image, which deals with the extracting statistical measures from this matrix. The Equation depicts the following

$$L=d(F(p)) \tag{1}$$

where L indicates the a class label (L), p is the pixel in the image, $F(p)$ the features of variable p, $d(F(p))$ is the decision function which maps the features of class label.

3.2.2 Statistical Attribute:

The statistical attribute deals with the inference of the major work or a task which is achieved with evaluation. It is used to interpret the data from the collection of data. It can be called as collective of facts. It should be enumerated and estimated accurately. The statistical approaches deal s with the context of various types of parameters which is accepted universally. The main parameters which are used under statistical attribute under image processing are mean, mode, median, variance, standard deviations, covariance, skewness and kurtosis. The Gray Level Co-occurrence Matrix (GCLM) is based on extraction of gray scale image [5] [17]. The main advantage of GLCM is to characterize the texture of the images and it calculates about the pair of the pixels with their corresponding values and it depends upon the spatial relationship that occurs in an image.

An image with the pixel values of paired pixel determined by

$$C_{\Delta x\Delta y}(i,j)=\sum_{x=1}^{n}\sum_{y=1}^{m}\begin{cases}1 & if\ I(x,y)=i\ and\ I(x+\Delta x,y+\Delta y)=j\\ 0 & otherwise\end{cases} \tag{2}$$

where the pixel values are represented by i and j, the spatial positions are represented by x and y for the image the offsets are (Δx Δy) defines the spatial relation for the matrix that is calculated. The image labeled by the cluster index is represented below.

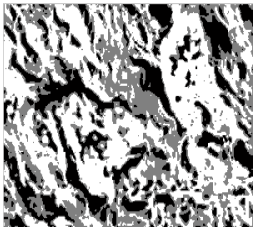

Fig.6. Image Labeled by Cluster Index using GCLM

3.2.3 Geometrical Attribute:

The geometrical attribute mainly deal with the projection of an image. The geometrical attribute is one of the feature extraction method. In image processing the features are extracted based on edge features, blobs, ridges, salient points , images textures and so on [16].Geometric features have proven to be very successful in applications such as object recognition, stereo matching, image retrieval, robot localization, video data mining, building panoramas, and recognition of object categories. The geometrical attribute extracts the features of an object (image) in various levels, centroid, area, perimeter, circularity and centre midpoint.

3.2.4 PCA Reduction:

The biomedical images are processed under efficient ways images are analyzed to provide the best functionality or effective way of determining the information. Hence rather than tables human recognize images for analyzing even though different methods are available for images to process one among them and proved to be the best is the PCA. The Principal Component Analysis is used for the reduction of dimensionality of the data [13]. If the images are taken under the low dimension to improve accuracy PCA is considered as an efficient method based on statistical representation of a random variable. PCA reduction is used under DWT.

Peak Signal to Noise Ratio (PSNR): PSNR is expressed in the form logarithmic scale.

$$PSNR = 10\times\log_{10}(255/\mathrm{RMSE}) \quad (3)$$

where the *PSNR* is calculated for the similarity between the referenced image and fused image.

The Root Mean Square error between the referenced images. High PSNR is said to be the better fused images.

Entropy (*E*): Entropy calculates the amount of information in the image:

$$E = \sum_{j=0}^{i-1} P_j \log_{10} P_j \quad (4)$$

The high entropy values referred to the increased information and improved fusion.

3.3 HIGH SUPPORT VECTOR MACHINE (HSVM)

The HSVM which is machine learning approaches which is considered to be the most contemporary approaches which is used for describing the predictions, identifications and of course making the decisions without the human involvement. The high support vector machine (HSVM) is a classification technique which provides satisfactory results in various applications. The image segmentation and classification are the main technique used under medical image Processing. The system is trained with the images by any type of pixel that belongs to the virus image. The images of cytomegalovirus is taken which is affected by various circumstances which is trained and tested. To classify multiple classes [10]. The associated algorithms deals with the learning algorithms which analyze data used for the classification and regression analysis [6] [7] [8]. The support vector machine is mainly focused to perform the training the example of that are most difficult to classify. The object detection is done using SVM to perform the accuracy [9]. The main aim is to provide the training and testing of the cytomegalovirus image with high quality of the image without any loss of information. This provides a better way to analyze and identify the virus indication or the impact of the virus in the human body when this work is incorporated with few medical additions based on their own technologies will provide a solution to the earlier detection of the virus which in turn provides an opportunity to overcome the complications in congenital infections.

4. EXPERIMENTAL RESULTS

4.1 CONFUSION MATRIX

The recital study is done based on the confusion matrix where the matrix shows the incorrect and correct predictions by the model which performs the actual outcome [11] [12]. Testing and validating or training and testing the dataset with the expected outcome values. The number of predictions which are correct is done with accuracy. The confusion matrix is used to designate the classification model. The terminology indicates a type of matrix consisting of contingency table with the actual and predicted two dimensions. It represents true positives, true negatives, false positives and false negatives.

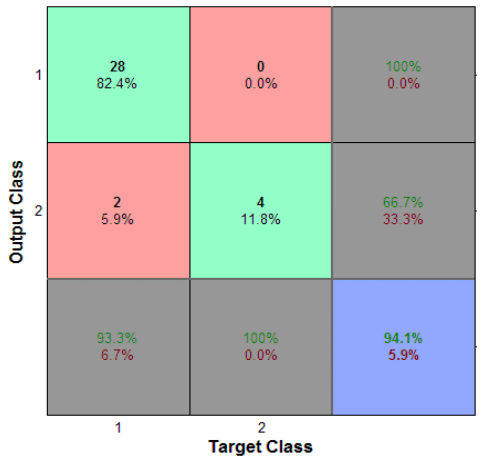

Fig.7. Confusion Matrix

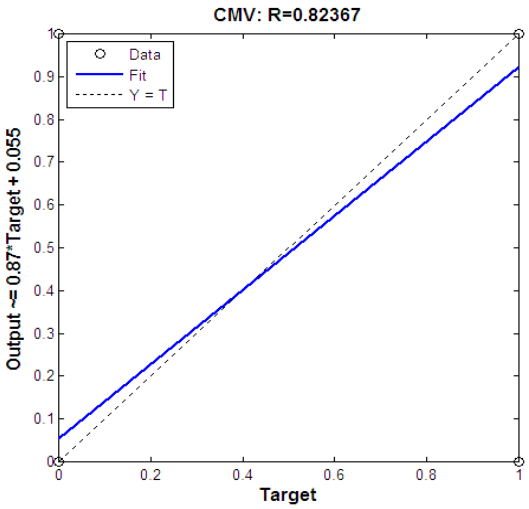

Fig.8. Graph of CMV Target Class

The CMV Target Class is corresponds to the confusion matrix to plot the rows correspond to the predicted class and the columns correspond to the true class (the Target class)

4.2 ACCURACY

The main aim is to provide the performance measures based on Accuracy. The basic idea is to put forward the diagnostic test interpretation and to calculate the test possibility of True Positive, True Negative, False Positive and False Negative.

$$Accuracy = (TP+TN)/(TP+TN+FP+FN) \quad (5)$$

4.3 PRECISION

The precision deals with the instances which are relevant and the retrieved instances are concentrated. The precision deals with the True Positive and False Positive to classify the possibility of the image clarity

$$\text{Precision}=TP/(TP+FP) \quad (6)$$

4.4 SPECIFICITY

The specificity measures the negatives which are correctly recognized. The specificity has the probability of finding the True Negative and False positive

$$Specificity = TN/(TN+FP) \quad (7)$$

4.5 SENSITIVITY

The sensitivity measures the positives that are correctly recognized. The sensitivity has the probability of True Positive and False Negative to prove the best results.

$$\text{Sensitivity}=TP/(TP+FN) \quad (8)$$

4.6 FALSE POSITIVE RATE (FPR)

False Positive Rate has the false positive and True Negative. This mainly concentrates on a false alarm will be raised the positive value is given when the true value is negative.

$$FPR=FP/(FP+TN) \quad (9)$$

4.7 F1 SCORE

*F*1-score is the measure of accuracy, the precision and recall the number of positive results. It deals with false positives and false negatives.

$$F1 = (2 \times Precision \times Recall)/(Precision + Recall) \quad (10)$$

4.8 MATHEWS CORRELATION COEFFICIENT (MCC)

The coefficient takes true and false positives, negatives, which provides the balanced measures even the classes are of different types.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (11)$$

4.9 KAPPA

The Kappa is identical to accuracy. It measures inter rater reliability of chance of agreement. P_o is observed agreement and P_e is hypothetical probability of chance agreement.

$$K = (P_o-P_e)/(1-P_e) \quad (12)$$

4.10 GEOMETRIC MEAN

The Geometric mean is defined as the n^{th} root of positive numbers.

$$\prod_{i=1}^{N} x_i = \sqrt[n]{a_1, a_2, ..., a_n} \quad (13)$$

where Π represents the product (images).

Table.1. Performance Measures of CMV

Performance Measures	Values
Accuracy	91.1765
Error	8.8235
Sensitivity	92.8261
Specificity	97.4713
Precision	86.6667
FP Rate	2.5287
F1 Score	88.8636
Matthews Correlation Coefficient	85.8323
Kappa	76.4706
Geometric Mean	85.1203

The performance measures of CMV is depicted in the image p_r show the parameters which finds the results with the maximum output is proved and the difference between the normal, partially affected, full affected are classified with the best results.

5. CONCLUSION

The image of cytomegalovirus is taken as input and the image is processed. The input image is taken and processed to form an intensified image using the image background subtraction and hierarchical ranking Convolutional Neural Network (HR-CNN) is used to perform feature extraction and segmentation of image is done using K-Means clustering and discrete wavelet transform to form an ROI image to have a better clarity of image without the loss of information. To make it more precise the image of cytomegalovirus is further processed with keen notification of the regions of interest using the Multiple Attribute Feature Extraction (MAFE) algorithm to show further clarity of the image using Principle Component Analysis reduction method is used to find the variance and co-variance to overcome low dimensional representation is maximized. The images are trained and tested based on HSVM to identify whether the images consists of partially affected, fully affected, uninfected. The awareness to be created among the people and the tests for the tests for the immunoglobulins can be made mandatory for the pregnant women as it is done for the thyroid, diabetes using TORCH test. So we could take some precautions to avoid the cytomegalovirus since it shows more impact on the newborns. Congenital infections could be avoided if these tests are done prior. At most care should be taken to prevent this kind of viruses by making ourselves to build good immune system.

REFERENCES

[1] Martin Ryner, Jan Olov Stromberg and Cecilia Soderberg, "Identification and Classification of Human Cytomegalovirus Capsids in Textured Electron Micrographs using Deformed Template Matching", *Virology Journal*, Vol. 3, pp. 1-18, 2006.

[2] C.C. Hung, E. Song and Y. Lan, "Image Texture, Texture Features, and Image Texture Classification", *Proceedings of International Conference on Image Texture Analysis*, pp. 3-14, 2019.

[3] William Henry Nailon, "Texture Analysis Methods for Medical Image Characterisation", Master Thesis, Department of Oncology Physics, Edinburgh Cancer Centre and School of Engineering, University of Edinburgh, pp. 1-122, 2016.

[4] Adrien Depeursinge, "Multiscale and Multidirectional Biomedical Texture Analysis", *Proceedings of International Conference on Biomedical Texture Analysis*, pp. 231-236, 2017.

[5] J.D.M. Roman, S.G. Guerroro and M.G. Torres, "Predictive Models for the Medical Diagnosis of Dengue: A Case Study in Paraguay", *Computational and Mathematical Methods in Medicine*, Vol. 2019, pp. 1-7, 2019.

[6] Arun Kumar, "Brain Tumor Classification using Hybrid Model of PSO And SVM Classifier", *Proceedings of International Conference on Advances in Computing, Communication Control and Networking*, pp. 1-7, 2018.

[7] A. Vidhyalakshmi and C. Priya, "A Study on Supervised Learning in Medical Image Grading using IoT", *International Journal of Recent Technology and Engineering*, Vol. 7, No. 5, pp. 1-14, 2019.

[8] Yingjie Tian, Yong Shi and Xiaohui Liu, "Recent Advances on Support Vector Machines Research, *Technological and Economic development Of Economy*, Vol. 18, No. 1, pp. 5-33, 2012.

[9] M. Durgadevi, "Performance Analysis of Classification Approaches for the Prediction of Type II Diabetes", *Proceedings of International Conference on Advanced Computing*, pp. 1-12, 2017.

[10] Mohamed Adel, Ahmed Kotb, Omar Farag, M. Saeed Darweesh and Hassan Mostafa, "Breast Cancer Diagnosis using Image Processing and Machine Learning for Elastography Images", *Proceedings of International Conference on Modern Circuits and Systems Technologies*, pp. 1022-1029, 2019.

[11] Usma Niyaz, "Advances in Deep Learning Techniques for Medical Image Analysis", *Proceedings of International Conference on Parallel, Distributed and Grid Computing*, pp. 20-22, 2018.

[12] Anne Humeau-Heurtier, "Texture Feature Extraction Methods: A Survey", *IEEE Access*, Vol. 7, pp. 8975-9000, 2019.

[13] Steffi Agino Priyanka, Yuan-Kai Wang and Shih Yu Huang, "Low-Light Image Enhancement by Principal Component Analysis", *IEEE Access*, Vol. 7, pp. 3082-3092, 2018.

[14] K.S. Bhagat and J.P. Chaudhari, "Image Segmentation and Classification for Medical Image Processing", *International Journal on Future Revolution in Computer Science and Communication Engineering*, Vol. 5, no. 1, pp. 45-52, 2017.

[15] Neelima Bagri and Punit Kumar Johari, "A Comparative Study on Feature Extraction using Texture and Shape for Content Based Image Retrieval", *International Journal of Advanced Science and Technology*, Vol. 80, No. 2, pp. 41-52, 2015.

[16] Sanket Rege, Rajendra Memane, Mihir Phatak and Parag Agarwal, "2D Geometric Shape and Color Recognition using Digital Image Processing", *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, Vol. 2, No. 6, pp. 1-14, 2013.

[17] Ni Wayan Surya Wardhani and Prayudi Lestantyo, "Cross-Validation Metrics for Evaluating Classification Performance on Imbalanced Data", *Proceedings of International Conference on Computer, Control, Informatics and its Applications*, pp. 1-14, 2019.

[18] Luay Fraiwan and Mohanad Alkhodari, "Neonatal Sleep Stage Identification using Long Short-Term Memory Learning System", *Proceedings of International Conference on Medical and Biological Engineering*, pp. 1-14, 2020.

[19] N. Sharmila Banu and S. Suganya, "Feature Selection and Classification for Detection of Seizures from EEG Signal", *Proceedings of International Conference on Emerging Trends in Information Technology and Engineering*, pp. 1236-1242, 2020.

[20] National CMV Foundation, Available at: https://www.nationalcmv.org/.